

# Package ‘KsPlot’

July 2, 2014

**Type** Package

**Title** Check the power of several statistical models.

**Version** 1.3

**Date** 2011-09-19

**Author** Issei Kurahashi

**Maintainer** Issei Kurahashi <i.kurahashi@gmail.com>

**Depends** e1071, caret, nnet, MASS, caTools, mda, glmnet, randomForest, mvpart

**Description** Check the power of a statistical model moving sample size k using half-Cross Validation.

**License** GNU General Public License Version 2

**LazyLoad** yes

**Repository** CRAN

**Date/Publication** 2012-10-29 08:57:13

**NeedsCompilation** no

## R topics documented:

|                       |          |
|-----------------------|----------|
| KsamplePlot . . . . . | 2        |
| Ksplot . . . . .      | 3        |
| <b>Index</b>          | <b>5</b> |

---

KsamplePlot

*Check the power of a statistical model.*


---

### Description

Check the power of a statistical model moving sample size k.

### Usage

```
KsamplePlot(X, y, Ksample = c(seq(40, 100, 10), 150, seq(200, 1000, 100)),
  Method = "lm", Caret = "No", size = 5, Type = "numeric", NCV = 5)
```

### Arguments

|         |  |
|---------|--|
| X       | Explanatory variables  |
| y       | Dependent variable   |
| Ksample | Numbers of sample size   |
| Method  | Method of how to predict the dependent variable. When Type="numeric", "lm" is linear regression (default), "svm" is Support Vector Machine, "nn" is Newral Network, "rf" is Random Forest, "mars" is Multipe Adaptive Regression Splines, "cart" is Classification and Regression Tree and "lasso" is LASSO. When Type="binary", "lm" is discriminant analysis (default), "logistic" is logistic regression, "cart" is Classification and Regression Tree, "svm" is Support Vector Machine, "nn" is Newral Network, "rf" is Random Forest. |
| Caret   | If Caret == "Yes", you can specify the Method caret form.  |
| size    | NN's node size   |
| Type    | Type = "numeric" or Type = "binary" when the dependent variable is continuous or binary.   |
| NCV     | Performation number of harlf cross-validation (harlf sample is training data and the other is test data).  |

### Value

A list containing entries:

|        |   |
|--------|---|
| Par    | Parameters estemated in one-phase exponential decay model of Ksplot() function. |
| Result | Results of 1 - exprained variance (Type="numeric") or 1 - AUC (Type="binary").  |
| Pred   | Raw predicted values of each Cross-Vadridation.                                 |

### References

Kurahashi, I. (2011) Model Exploration using K-sample Plot in Big Data. <http://isseing333.blogspot.com/2011/10/model-exploration-using-k-sample-plot.html>

**Examples**

```

library(KsPlot)

set.seed(1)
x1 <- rnorm(1000000)
set.seed(2)
x2 <- rnorm(1000000)
set.seed(3)
y <- 2*x1 + x2**2 + rnorm(1000000)

X1 <- data.frame(x1 = x1, x2 = x2)
X2 <- data.frame(x1 = x1, x2 = x2, x3 = x2**2)
y <- y

# numeric outcome
set.seed(1)
KsResult1 <- KsamplePlot(X1, y)
set.seed(1)
KsResult1 <- KsamplePlot(X1, y, NCV=10)
set.seed(1)
KsResult2 <- KsamplePlot(X1, y, Method="svm")
set.seed(1)
KsResult3 <- KsamplePlot(X2, y)

# binary outcome
y1 <- ifelse(0 < y, 1, 0)

set.seed(1)
KsResult4 <- KsamplePlot(X2, y1, Type="binary")
KsResult4$Pred[1:100, ]

# random forest is overfitting to training sample but also have good response to test sample.
set.seed(1)
KsResult5 <- KsamplePlot(X2, y1, Type="binary", Method="rf")

```

---

Ksplot

*Draw K-sample plot.*


---

**Description**

Draw K's plot for function KsamplePlot using KsamplePlot object.

**Usage**

```
Ksplot(UseDataTest, UseDataTrain, Type = "numeric", Method)
```

**Arguments**

|              |  |
|--------------|--|
| UseDataTest  | KsamplePlot object of test data.   |
| UseDataTrain | KsamplePlot object of training data.   |
| Type         | Type = "numeric" or Type = "binary" when the dependent variable is continuous or binary. |
| Method       | The method specified in KsamplePlot.   |

# Index

KsamplePlot, 2  
Ksplot, 3